Working Group on the Future of Bibliographic Control: Economics and Organization of Bibliographic Data

Rick Lugg, R2 Consulting LLC
July 9, 2007

Economics of Bibliographic Control

In my formative years in this profession, I worked for a book vendor/library supplier. We saw ourselves as a company with a mission, and cultivated on behalf of our library customers an ethic of timely service and rapid response – an atmosphere of intensity and urgency. Occasionally, some of us would lose perspective, and begin to approach panic if a shipment or a set of cataloging records threatened not to be delivered on time. I had (and sometimes still have) a completely undeserved reputation as a wise guy. I became fond of reminding everyone that "there's no such thing as a bibliographic emergency." No one's life or health was hanging on our actions. This is not a justification for slacking, but please, let's just calm down and get back to work. To help our customers retain perspective, I even added the phrase to my out-of-office voice mail message: "Should you experience a bibliographic emergency, please call …"

Well, it turns out that karma is a wiseguy, too. Owing to the natural perversity of the universe, my partner Ruth Fischer and I now spend most of our time as, basically, bibliographic trauma specialists. Our firm does analysis and redesign of library workflows. We are seldom called in to admire how well things are going.

Instead, we are escorted around the library as if it were an accident scene: we view backlogs, frontlogs, and working queues. We are shown, with barely suppressed horror and survivors' glee, the unfinished retrospective conversions, the Dewey Collection that can't be moved to storage because it's not barcoded, the aftermath of ILS migrations, failed match points, the massive East Asian gift collection accepted by the Director, and the unacceptable records created by "other" libraries.

Some of these scenes are spectacular: cataloging backlogs with their own inventory systems and warehouses; rooms bursting with sagging boxes of gifts, many obviously inappropriate; Special Collections offices that would make OSHA shudder. But at least in the print world the problems have the good manners to be visible.

In the digital world, backlogs are more insidious; because they are invisible, and don't take up space, they can sometimes be forgotten or ignored. Where are those 100,000 bib records that came with EEBO, anyway? Where do we stack the holdings records awaiting "reclamation?" Where's the forklift for the data needed to populate that new ERMS. That automated link checker spews out broken links constantly—has anyone seen them? And where do we keep that blasted metadata everyone wants? On the surface, things often look tidier and more under control than they actually are. The undercurrent of tension, though,

is in some ways worse, because no one is ever completely sure how much work is lurking out there.

But whether the material in question is print or electronic, most libraries are struggling to cope with creating and sustaining access to their growing riches. Granted, our view may be skewed by always working where there are acknowledged problems. But my sense, from listening to others, from reading the literature, the listservs and the blogs (and the columns), is that a sense of "long emergency" is commonplace, and can be succinctly stated like this:

- Libraries cannot afford to support the level of bibliographic control they believe they should provide.

- The costs of traditional cataloging and bibliographic control have begun to seem disproportionate to their benefits.

- We need to take better advantage of the work done and the value added by other participants in the bibliographic "supply chain."

COSTS AND BENEFITS

Of course, these are not new problems. It's always been hard to keep up with the making of books, and even more so serials. Backlogs have been a part of libraries since long before any of us entered the profession. But cost issues have been thrown into relief much more dramatically by an explosion of newly available content, new discovery tools, and by the Web's culture of distributed

power and self-service. For the first time in history, libraries have competitors in the realm of search, and some of them are very good indeed. What's more, they are not bound by the same rules. Worse, this competition has intensified at a time when libraries face continuing pressure on staff budgets, especially for technical services operations. Finding or creating the capacity to bring the unique strengths of libraries to the Web is a management challenge of the first order.

It's not that libraries haven't done some grappling with costs. In many ways, they have spent the last three decades trying to control and reduce the cost of cataloging. They have relied on The Library of Congress and PCC to produce definitive cataloging records; on the cooperative spirit that shaped the growth of OCLC; on the emergence of vendor-supplied records (first in public libraries, then in academics); the growth of consortia and shared catalogs; and the creation of new approaches to discovery—link resolvers, A-Z lists, e-content knowledgebases. In short, libraries have relied on many of the "stakeholders" gathered here today for help in controlling costs.

To a large degree this effort has succeeded. In some respects, the basic MARC record – for English-language monographs, but also increasingly for e-journals-- has become a commodity. At its best, the principle of widely-shared effort— catalog once, distribute and use widely—has succeeded wildly. Even if it costs, as some have estimated, $150-200 to fully catalog an English-language

monograph, including all the necessary authority work, the use of that record in hundreds of libraries distributes the cost very reasonably. Those efforts have served users well.

Other cost-control efforts have been instituted or at least attempted by libraries, some very recently, such as the "Access Level record for Serials". Non-MLS copy catalogers and FastCat operations, which combine Acqusitions and Cataloging, are now common. It is not unusual to see paraprofessional staff doing some level of original cataloging. Nonetheless, even after 30 years of shared cataloging and efforts like these, there are still staggering duplicative costs in the system. What makes some of these costs so persistent?

BARRIERS TO COST REDUCTION

Every day, in libraries all over the country, perfectly good LC records are subjected to all manner of scrutiny and revision. 100s and 245s and 260s and 490s are checked, to see if LC got them right. The 300 field draws special attention, as books are re-measured and pagination double-checked. Call numbers are examined, and compared against the local catalog to eliminate duplication. Custom Cuttering schemes are applied. Call numbers are written in pencil on the verso. Flags of every color are inserted.

At the root of these processes are two powerful beliefs. One: the cult of perfection. And two: cataloging is about how print books are arranged on the shelf. Let's look at them quickly in turn.

The cult of perfection has its positive attributes; accuracy and detail are important elements in bibliographic control. In its darker aspect, however, it leads to a focus on the record and the rules for their own sake rather than its actual use. The "artisan" approach to bibliographic records can easily be overdone, and can lead to a narrow view of quality—one which does not consider timeliness or elements of equal importance to users. There will always be errors; cataloging is a human activity. The quality question should not be "is this record perfect?" but rather "Does this error impede access?" The cost question also pertains here: how much effort can we afford to expend on this one record, when 1,000 more titles are waiting? In the realm of original cataloging, is there in fact any record (with rare exceptions) that is worth a full day of a professional cataloger's time—a figure we have heard more than once. Can libraries really afford this level of attention to detail? And even if they can, are there users who actually require it?

Not surprisingly, shelf-arrangement retains a powerful influence on cataloging; the call number has historically been an address. While this still has some significance, its importance is diminishing as more electronic material is available, as more print material resides in remote storage, or automated retrieval centers browseable only by robots. Duplicate call numbers, while technically incorrect, don't typically prevent access. And while collocation of artists or composers or literary authors can be very helpful, that browsing experience can occur via a system interface rather than by relying on

adjustment of individual records. We need to transfer our thinking about collocation and links to full-text digital content, and invest less in the print-bound paradigm of shelf-arrangement.

NEW COSTS

The proliferation of electronic content has also created new costs, involving new kinds of records, systems, and access paths. Perhaps most significantly for our discussion, it has launched a new category of cost related to bibliographic control, as libraries choose whether to maintain a single record for print and electronic versions of journals or books—or whether to tolerate multiple records that may be more easily updatable.

TOTAL COST OF CATALOGING

The initial efforts associated with subject analysis and classification represent only one portion of the investment in bibliographic control. As with other investments, it's useful to consider the full life-cycle costs of the process—from selection through digitization or withdrawal. Although the content does not change, its form and representation in the catalog often do. For serials, there is constant maintenance of check-in records, holdings records, and title changes as titles make their way through binding, offsite storage, replacement with microfilm or electronic versions. Shifting, removal to automated retrieval centers or offsite storage, "de-ref" ing, withdrawals—all of these changes involve item-by-item record maintenance. Even with batch processes and automation

support, subsequent database management can far outpace the cost of initial cataloging. The full costs of bibliographic control are probably actually much higher than any estimates we've seen. Perhaps one of the biggest ironies we see in our work is that cancellation and reduction of print subscriptions actually increases the number of hours required to support print. Although this activity resembles inventory control more than bibliographic control, the maintenance of item records, holding records, check-in records absorbs enormous amounts of staff time that could support bibliographic control.

OPPORTUNITY COSTS

Any discussion of costs ultimately is a discussion of priorities: are we putting our effort where it can yield the most benefit? The concept of "opportunity cost" adds an important dimension to this question: what are we NOT able to do that we should be doing? What are the unsolved problems? Where do we need more capacity? These are the kinds of things we hear routinely:

- Special Collections cataloging: rare books, non-books, unique material, finding aids, EAD encoding—world-class backlogs.

- Support for Manuscripts and Archives: finding aids need to be created and encoded

- Theses & dissertations

- Involvement in non-MARC metadata

- Course management systems like Blackboard and Moodle—how do we deliver library resources into the undergraduate workflow?

- Mass digitization: what new retrieval opportunities will the existence of millions of full-text volumes create?

- Institutional repositories: earlier involvement from Cataloging is needed to advise on metadata standards, to consult on content description and access.

There is a cost to delay in these kinds of initiatives; the library can lose its place. These areas represent the library's future, if it is to remain vital and relevant. This is where more investment is needed, where the trade-offs that are so hard to make begin to pay off. Let's conclude with these questions:

- How do we reduce our efforts on commonly-available material? How can we take better advantage of work already done by publishers, LC, vendors, agents and other libraries?

- How do we redirect our focus to the most unique material—to create records that no other institution is likely to create?

- How do we find the capacity to shift our expertise to new formats and new metadata schemes more quickly and fully?

- How do we open our system, and more importantly, our cultures more fully to collaborate directly with authors, publishers, vendors, A&I services, and users?

These are some of the economic questions to be taken up here today; now, in our last few minutes, let's turn to questions of organization.

Organization of Bibliographic Control

The library is, of course, only one link in the supply chain of bibliographic

information between author and reader. Its needs are unique, but not

necessarily exclusive. All parties contribute value through the vehicle of the bib

record: Creator, Publisher, Vendors/Distributors, Stores/Libraries. To date, there is

not a strong tradition of sharing data and metadata throughout the publication

cycle. It may be useful, then, to think about what information is available at

each stage, and how to aggregate and build on that foundation. What value is

added at each stage? How can the existing value be captured and leveraged

in the next?

CREATOR
Author abstracts

PUBLISHER: EDITORIAL
catalog/marketing copy
The file or galley itself
ONiX records
CIP submission form
Subject focus
Publication history,
Rights information
Author information

PUBLISHER: PRODUCTION:
Indexes
Tables of Contents
DOI @ chapter or article-level
Mark-up with XML tags
Conversion
Full-text

MARKETING/DISTRIBUTION
Vendor descriptors/approval plan apparatus: an entire businesses developed around things that LC MARC record does NOT do well: (limitations of LCSH and LC Class): topical, format, interdisciplinary descriptors
Preliminary classification (assigned by buyers)
Advance reviews
Full-text (eBook version/articles/ pre-prints)
A&I Services
BIC/BISAC descriptors for retail
Amazon "look inside the book"
Publisher flap copy; author bio
Cover scan
Knowledgebases
Cataloging records/record sets
          EEBO, ECCO, netLibrary, ebrary,
          Juvenile record
          Foreign-language

AT POINT OF SALE/CIRCULATION:
Libraries: What value do libraries add, and what does it cost? Whom does it serve?
          Classification (LC, DDC, NLM)
          Controlled vocabulary subject headings/authority
          Name authority
          Series authority (?)
          Other records: Bib, holdings, item…consortial, URLs

          What elements are ignored or removed

How do we get the best of it all on behalf of the user? The controlled, reliable

MARC structure that makes an Endeca front-end so rich; the user generated

tags that enlist the help of users; The "loosely federated" record would

accumulate the value of all parties, allowing everyone contribute. Perhaps

instead of insisting on our centrality and importance in the chain, we need to

find and focus on our place in it.